Annotations *Rhapsodie*

pour le *Trameur*

13/12/2013 12:51:24

Serge Fleury

Références

Le Trameur, manuel d'utilisation

http://www.tal.univ-paris3.fr/trameur/leMetierLexicometrique.pdf

Dans cette documentation, la partie « *Relations de dépendance entre les items de Trame (via leurs annotations)* » présente l'état des développements actuellement disponibles dans le *Trameur* pour l'exploitation des relations de dépendance.

Le Trameur. Propositions de description et d'implémentation des objets textométriques

http://www.tal.univ-paris3.fr/trameur/trameur-propositions-definitions-objets-textometriques.pdf

Ce document met au jour une description des objets textométriques et les méthodes mises en œuvre dans le ***Trameur*** pour travailler sur et avec ces objets dans une perspective textométrique. On y détaille aussi les opérations permises sur une *base textométrique* : format des données textuelles, modification dynamique de la *Trame*, correction ou ajout d'annotation etc.

Sommaire

1. Préambule

Ce document commence par décrire le processus de transcodage des données issues du projet *Rhapsodie* (http://projet-rhapsodie.fr/) sous la forme d'une base textométrique importable dans le *Trameur*.

Il présente ensuite les différentes fonctionnalités mises en œuvre pour traiter ce type de données via le *Trameur*.

2. Données Rhapsodie

Les données traitées sont disponibles sur la page du projet *Rhapsodie* :

- Téléchargement des fichiers de codage microsyntaxique version béta 10/13 (zip)
- Tutoriel codage microsyntaxique (pdf)

Le fichier d'annotations (`Rhapsodie.tok`) a l'allure suivante (lecture ici dans un tableur) :

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	TextID	TreeID	TokenID	Token	Lemma	POS	Mode	Tense	Person	Number	Gender	Gov_rection	Type_rection	Gov_para	Type_para	Gov_inher	Type_inher	Gov_junc	Type_junc	Gov_junc-inher	Type_junc-inher
2	M2006	1	1	bonjour	bonjour	B_J						0	root								
3	M2006	1	2																		
4	M2006	1	3	Eric	Eric	B_N				sg	masc	0	root								
5	M2006																				
6	M2006	2	1	bonjour	bonjour	B_N				sg	masc	0	root								
7	M2006	2	2																		
8	M2006	2	3	Ã	Ã	B_Pre						1	dep								
9	M2006	2	4																		
10	M2006	2	5	tous	tous	B_Pro			3	pl	masc	3	dep								
11	M2006																				
12	M2006	3	1	nouvelle	nouveau	B_Adj				sg	fem	3	dep								
13	M2006	3	2																		
14	M2006	3	3	nuit	nuit	B_N				sg	fem	0	root								
15	M2006	3	4																		
16	M2006	3	5	de	de	B_Pre						3	dep								
17	M2006	3	6																		
18	M2006	3	7	pillage	pillage	B_N				sg	masc	5	dep								
19	M2006	3	8																		
20	M2006	3	9	et	et	B_J												5	junc		
21	M2006	3	10																		
22	M2006	3	11	d	de	B_Pre								5	para_coord	3	dep_inherited	9	junc		
23	M2006	3	12	'	^	I_Pre															
24	M2006	3	13	affrontement	affrontement	B_N				sg	masc	11	dep								
25	M2006	3	14																		
26	M2006	3	15	en	en	B_Pre						3	dep								
27	M2006	3	16																		
28	M2006	3	17	Guadeloupe	Guadeloupe	B_N				sg	masc/fem	15	dep								
29	M2006																				
30	M2006	4	1	trois	trois	B_D				sg	masc	3	dep								
31	M2006	4	2																		
32	M2006	4	3	policiers	policier	B_N				pl	masc	0	root								
33	M2006	4	4																		
34	M2006	4	5	blessÃ©s	blessÃ©	B_Adj				pl	masc	3	dep								
35	M2006																				

Ces données sont constitués par un certain nombre de textes (l'identifiant du texte est visible dans la première colonne), chacun d'eux est segmenté en « unité illocutoire » (UI, seconde colonne), chacune d'elle est segmentée en `token` (troisième colonne), chacun d'eux est annoté (les autres colonnes)

Le fichier tabulé précédent est composé de 21 colonnes (description à compléter cf *Rhapsodie*) :

1	TextID	Identifiant de l'échantillon (nom de la PARTIE dans la terminologie textométrique)
2	TreeID	Numéro de l'unité illocutoire (UI) dans l'échantillon
3	TokenID	Identifiant du *token* dans l'UI Les UI d'un échantillon sont séparées les unes des autres par des lignes sans aucun identifiant *TreeID*
4	Token	Segment de la transcription orthographique pris en 2 blancs ou un blanc et une signe de ponctuation
5	Lemma	Les lemmes sont comme il est d'usage la forme pour les lexèmes invariables, la forme infinitive pour les verbes, le singulier pour les noms et le masculin singulier pour les adjectifs.
6	POS	Partie du discours - V pour les verbes - N pour les noms - Adj pour les adjectifs - Adv pour les adverbes - Pre pour les prépositions - CS pour les conjonctions de subordination - J pour les joncteurs : il s'agit des traditionnelles conjonctions de coordinations et d'autres éléments qui lient les couches d'un entassement, comme *c'est-à-dire* ou *y compris*. Les éléments clôtureurs d'entassement comme *et caetera* sont classés comme joncteurs également. - D pour les déterminants - I pour les interjections, y compris des marqueurs de discours comme *bon, ben, euh, hein* ... - Qu pour les mots qu- que sont les relatifs et les interrogatifs - Cl pour les clitiques, y compris les clitiques sujets (*je, tu, il, on, ce*) et l'adverbe de négation *ne*. - Pro pour les autres pronoms - X pour les éléments dont on ne peut déterminer la catégorie syntaxique : partie inaudible (XXX), certaines amorces (quand on ne peut pas deviner le lexème et sa partie du discours), ainsi que les positions non instanciées marquées par &.
7	Mode	Les V reçoivent un trait de mode qui peut prendre 6 valeurs : *indicative, subjunctive, imperative, infinitive, past_participle, present_participle*
8	Tense	Seuls les V à l'indicatif varient en temps ; le trait *tense* possède 5 valeurs : *present, imperfect, future, conditional* et *perfect*
9	Person	Les V reçoivent aussi des traits d'accord : le trait *person* a trois valeurs *1, 2* et *3*
10	Number	le trait *number* a deux valeurs *sg* et *pl*
11	Gender	le trait *genre* a deux valeurs *fem* et *masc*
12	Gov_rection	
13	Type_rection	
14	Gov_para	
15	Type_para	
16	Gov_inher	
17	Type_inher	
18	Gov_junc	
19	Type_junc	
20	Gov_junc-inher	
21	Type_junc-inher	

Les 11 premières annotations sont réutilisées telles quelles par le processus de transcodage.

Les suivantes sont réutilisées 2 à 2 (*cf* jeu de couleur) pour construire respectivement une seule annotation (de type relation) qui est réécrite par exemple sous la forme : `Type_rection(Gov-rection)` pour les lignes 12 et 13.

3. Intégration des annotations Rhapsodie dans une base textométrique

Les données issues de *Rhapsodie* transcodées dans un format compatible avec le *Trameur* respectent la structuration d'une base textométrique (*cf* documentation *Trameur*). Celle-ci est composée de 2 parties permettant :

1. La description d'une *Trame* textométrique : liste des items numérotés et annotés (ici chaque item est associé à 13 annotations)

2. La description du *Cadre* textométrique : liste des partitions définies sur la trame ; chacune porte un nom et est associée à une liste de parties définies chacune par son nom (le nom de l'échantillon de *Rhapsodie*), par sa position de début sur la *Trame* et sa position de fin

Le fichier issu du transcodage est au final une base textométrique qui se charge dans le *Trameur* par le module « *importation de base* ».

3.1 La Trame textométrique

Le fichier construit a l'allure suivante (on donne à voir les mêmes UI que ci-dessus). La *Trame* textométrique définit la liste des items qui la composent :


```
<items>
<item type="forme" pos="1"><f>bonjour</f><c>B_I</c><l>bonjour</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>ROOT</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="2"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="3"><f>Eric</f><c>B_N</c><l>Eric</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>ROOT</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="4"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="5"><f>$</f><c>DELIM</c><l>DELIM</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="6"><f>RETURN</f><c>RETURN</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="7"><f>bonjour</f><c>B_N</c><l>bonjour</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>ROOT</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="8"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="9"><f>à</f><c>B_Pre</c><l>à</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>DEP(7)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="10"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="11"><f>tous</f><c>B_Pro</c><l>tous</l><a>-</a><a>-</a><a>3</a><a>pl</a><a>masc</a><a>DEP(9)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="12"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="13"><f>$</f><c>DELIM</c><l>DELIM</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="14"><f>RETURN</f><c>RETURN</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="15"><f>nouvelle</f><c>B_Adj</c><l>nouveau</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>fem</a><a>DEP(17)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="16"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="17"><f>nuit</f><c>B_N</c><l>nuit</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>fem</a><a>ROOT</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="18"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="19"><f>de</f><c>B_Pre</c><l>de</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>DEP(17)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="20"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="21"><f>pillage</f><c>B_N</c><l>pillage</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>DEP(19)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="22"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="23"><f>et</f><c>B_J</c><l>et</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>JUNC(19)</a><a>-</a><a>-</a></item>
<item type="delim" pos="24"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="25"><f>d</f><c>B_Pre</c><l>de</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>PARA_COORD(19)</a><a>DEP_INHERITED(17)</a><a>JUNC(23)</a><a>-</a><a>-</a></item>
<item type="forme" pos="26"><f>'</f><c>I_Pre</c><l>^</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="27"><f>affrontement</f><c>B_N</c><l>affrontement</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>DEP(25)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="28"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="29"><f>en</f><c>B_Pre</c><l>en</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>DEP(17)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="30"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="31"><f>Guadeloupe</f><c>B_N</c><l>Guadeloupe</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc-fem</a><a>DEP(29)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="32"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="33"><f>$</f><c>DELIM</c><l>DELIM</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="34"><f>RETURN</f><c>RETURN</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="35"><f>trois</f><c>B_D</c><l>trois</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>DEP(37)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="36"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="37"><f>policiers</f><c>B_N</c><l>policier</l><a>-</a><a>-</a><a>-</a><a>pl</a><a>masc</a><a>ROOT</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="38"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="39"><f>blessés</f><c>B_Adj</c><l>blessé</l><a>-</a><a>-</a><a>-</a><a>pl</a><a>masc</a><a>DEP(37)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="40"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="41"><f>$</f><c>DELIM</c><l>DELIM</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="42"><f>RETURN</f><c>RETURN</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
```


Chaque token de *Rhapsodie* devient un *item* de la *Trame*, il est identifié par sa position sur la *Trame* : l'identification des tokens définie dans les relations de dépendance fournies au départ dans *Rhapsodie*, est donc, après transcodage, la position de l'item dans la base (l'identification n'est plus déterminée au sein de la UI, elle est définie par la position dans la base construite).


```
<item type="delim" pos="46"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="47"><f>lance</f><c>B_V</c><l>lancer</l><a>indicative</a><a>present</a><a>3</a><a>sg</a><a>-</a><a>ROOT</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="48"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="49"><f>u  OBJ  _D</c><l>un</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>DEP(51)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="50"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="forme" pos="51"><f>appel</f><c>B_N</c><l>appel</l><a>-</a><a>-</a><a>-</a><a>sg</a><a>masc</a><a>OBJ(47)</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="52"><f> </f><c>BLANK</c><l>BLANK</l><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a><a>-</a></item>
```


Toutes les annotations initiales sont associées à chacun des items de la *Trame*.

Les annotations de dépendance sont fusionnées deux à deux sous la forme d'une relation marquée de la manière suivante : RELATION(CIBLE) (*cf* documentation *Trameur*)

- RELATION est une chaîne portant le nom de la relation visée
- CIBLE est une valeur numérique pointant vers une position de la *Trame*

Chaque item est associé *in fine* à 13 annotations, la figure suivante donne à voir ces annotations sur un item en contexte :

nous ferons le point sur les violences en direct avec notre envoyé spécial Sébastien Paour dans quelques instants §

```
Position:<91>
Forme:<point>|Freq:34
Lemme:<point>|Freq:23
Cat:<B_N>|Freq:6247
a-00004:<->|Freq:70852
a-00005:<->|Freq:72497
a-00006:<->|Freq:68133
a-00007:<sg>|Freq:17353
a-00008:<masc>|Freq:8816
a-00009:<OBJ(87)>|Freq:1
a-00010:<->|Freq:74709
a-00011:<->|Freq:74145
a-00012:<->|Freq:75936
a-00013:<->|Freq:76812
```

n°Annotation	Label	Contenu
1	Forme	Forme graphique
2	Lemme	Lemme
3	Cat	P.O.S
4	a-00004	Mode
5	a-00005	Tense
6	a-00006	Person
7	a-00007	Number
8	a-00008	Gender
9	a-00009	Type_rection(Gov_rection)
10	a-00010	Type_para(Gov_para)
11	a-00011	Type_inher(Gov_inher)
12	a-00012	Type_junc(Gov_junc)
13	a-00013	Type_junc-inher(Gov_junc-inher)

3.2 Le Cadre textométrique

Les différents échantillons initiaux de *Rhapsodie* sont considérés comme autant de parties différentes : la base finale est donc une partition de textes (*Cadre*), chaque partie contient les zones textuelles associées à l'identifiant initial de l'échantillon. Ci-dessous, le *Cadre* final mis au jour dans le *Trameur* et son codage dans la base construite après transcodage :

Le Cadre Lexicométrique


```
<acces>
<partition nom="PARTIE">
 <p n="D0017" d="34004" f="34238" nd="47" nf="48"/>
 <p n="M0001" d="28501" f="28801" nd="33" nf="34"/>
 <p n="M0019" d="35358" f="35768" nd="51" nf="52"/>
 <p n="M0024" d="65439" f="65709" nd="97" nf="98"/>
 <p n="M1001" d="57263" f="58103" nd="81" nf="82"/>
 <p n="M2003" d="17835" f="19359" nd="25" nf="26"/>
 <p n="M0004" d="5501" f="5599" nd="9" nf="10"/>
 <p n="M0011" d="59602" f="59946" nd="87" nf="88"/>
 <p n="M0014" d="76677" f="76850" nd="113" nf="114"/>
 <p n="D1002" d="50286" f="51338" nd="71" nf="72"/>
 <p n="M2005" d="75179" f="75909" nd="109" nf="110"/>
 <p n="D2012" d="60955" f="63091" nd="93" nf="94"/>
 <p n="D1003" d="6075" f="7263" nd="13" nf="14"/>
 <p n="M1003" d="29045" f="30753" nd="37" nf="38"/>
 <p n="D2013" d="58233" f="59601" nd="85" nf="86"/>
 <p n="D2002" d="55230" f="57262" nd="79" nf="80"/>
 <p n="M0007" d="33310" f="33698" nd="43" nf="44"/>
 <p n="M2004" d="19360" f="21964" nd="27" nf="28"/>
 <p n="M0009" d="7264" f="7750" nd="15" nf="16"/>
 <p n="M0006" d="28802" f="29044" nd="35" nf="36"/>
 <p n="M0008" d="58104" f="58232" nd="83" nf="84"/>
 <p n="M0021" d="54855" f="55229" nd="77" nf="78"/>
 <p n="M2006" d="1" f="2083" nd="1" nf="2"/>
 <p n="D2004" d="26292" f="28500" nd="31" nf="32"/>
 <p n="M0018" d="60420" f="60954" nd="91" nf="92"/>
 <p n="D0006" d="44898" f="47922" nd="67" nf="68"/>
 <p n="D0005" d="35882" f="38116" nd="55" nf="56"/>
 <p n="D0007" d="10254" f="10566" nd="19" nf="20"/>
 <p n="M0015" d="40260" f="40424" nd="59" nf="60"/>
 <p n="D2005" d="71967" f="74131" nd="105" nf="106"/>
 <p n="M0002" d="54452" f="54854" nd="75" nf="76"/>
 <p n="D2003" d="66091" f="69203" nd="101" nf="102"/>
 <p n="D1001" d="38117" f="40259" nd="57" nf="58"/>
 <p n="D2007" d="30754" f="33168" nd="39" nf="40"/>
 <p n="D2008" d="13234" f="17834" nd="23" nf="24"/>
 <p n="D2001" d="21965" f="26291" nd="29" nf="30"/>
 <p n="D0001" d="10567" f="13233" nd="21" nf="22"/>
 <p n="D2010" d="7751" f="10253" nd="17" nf="18"/>
 <p n="M0016" d="4988" f="5500" nd="7" nf="8"/>
 <p n="D0020" d="44667" f="44897" nd="65" nf="66"/>
 <p n="D2009" d="2781" f="4987" nd="5" nf="6"/>
 <p n="D0004" d="41934" f="44666" nd="63" nf="64"/>
 <p n="D2011" d="63092" f="65438" nd="95" nf="96"/>
 <p n="M0013" d="65710" f="66090" nd="99" nf="100"/>
 <p n="D0002" d="47923" f="50285" nd="69" nf="70"/>
 <p n="D0009" d="51339" f="54451" nd="73" nf="74"/>
 <p n="D0003" d="74132" f="75178" nd="107" nf="108"/>
 <p n="M0022" d="59947" f="60419" nd="89" nf="90"/>
 <p n="M2002" d="69204" f="71966" nd="103" nf="104"/>
 <p n="M0012" d="35769" f="35881" nd="53" nf="54"/>
 <p n="M0003" d="75910" f="76676" nd="111" nf="112"/>
 <p n="D2006" d="5600" f="6074" nd="11" nf="12"/>
 <p n="M0010" d="33169" f="33309" nd="41" nf="42"/>
 <p n="D0008" d="40425" f="41933" nd="61" nf="62"/>
 <p n="M0023" d="2084" f="2780" nd="3" nf="4"/>
 <p n="M0005" d="33699" f="34003" nd="45" nf="46"/>
 <p n="M2001" d="34239" f="35357" nd="49" nf="50"/>
</partition>
</acces>
```

3.3 Sections

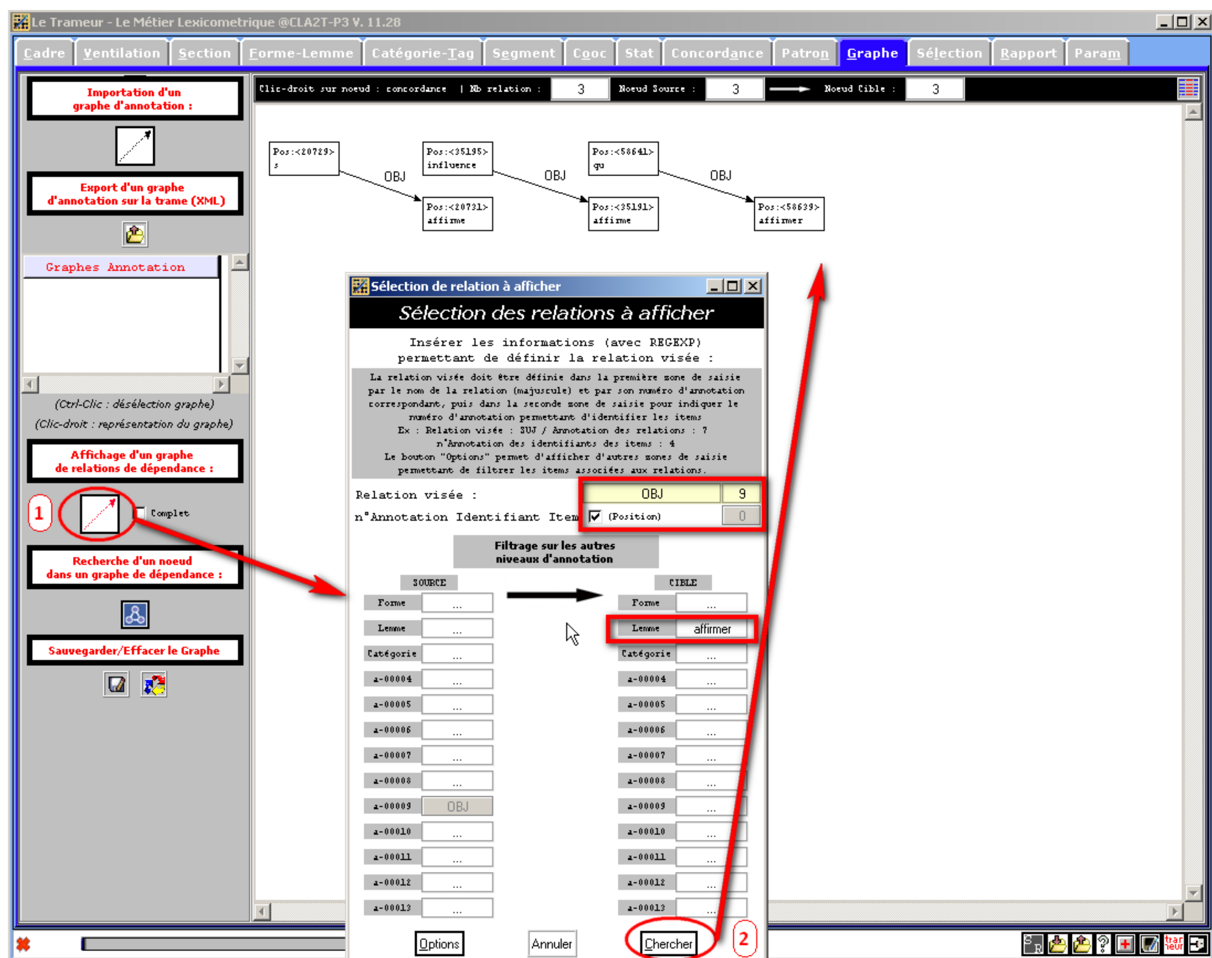
Le processus de transcodage intègre aussi un marquage de sections : après chaque UI, un caractère délimiteur de section (§) est introduit (*cf* caractère en position 5 dans la base présentée ci-dessus) pour permettre de construire dans le *Trameur* une représentation cartographique de la base sous la forme d'une *carte des sections* :

4. Explorer les relations de dépendance

Les différentes fonctionnalités disponibles dans le *Trameur* pour travailler avec les annotations de relations sont décrites dans la documentation du *Trameur*. On les illustre ci-dessous sur les données de la base *Rhapsodie2Trameur*.

4.1 Recherche de dépendance sur l'ensemble de la base (avec filtrage sur les items en relation)

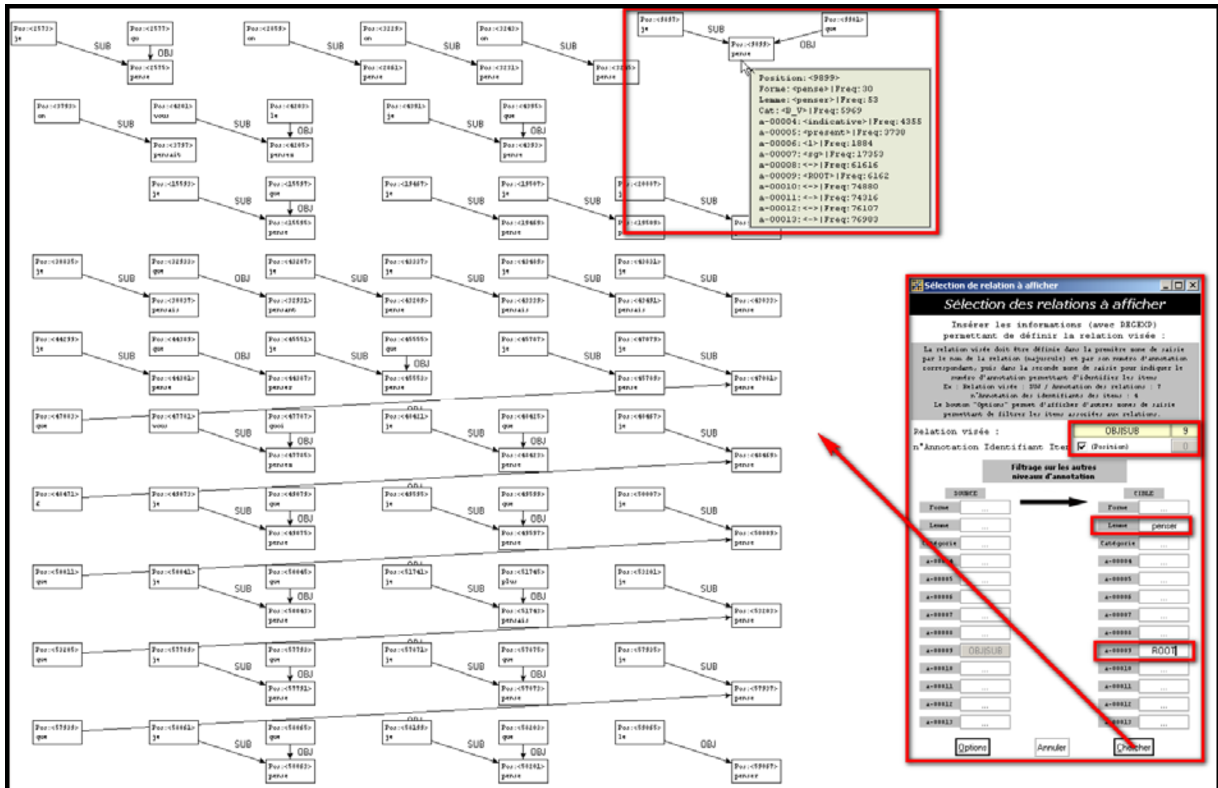
Exemple n°1 : recherche des « objets » du lemme « affirmer »



Dans la figure précédente, on cherche des relations de dépendance de type OBJ en contraignant un des composants de la relation : il doit porter le lemme "affirmer" (i.e. on cherche les objets d'affirmer).

Exemple n°2 : recherche des « sujets » et « objets » du lemme « penser »

Dans la figure suivante, la relation cherchée est double via l'expression régulière SUB|OBJ *i.e* SUB ou OBJ, la cible de la relation impose une valeur pour le lemme (« penser ») et pour l'annotation n°9 (ROOT)

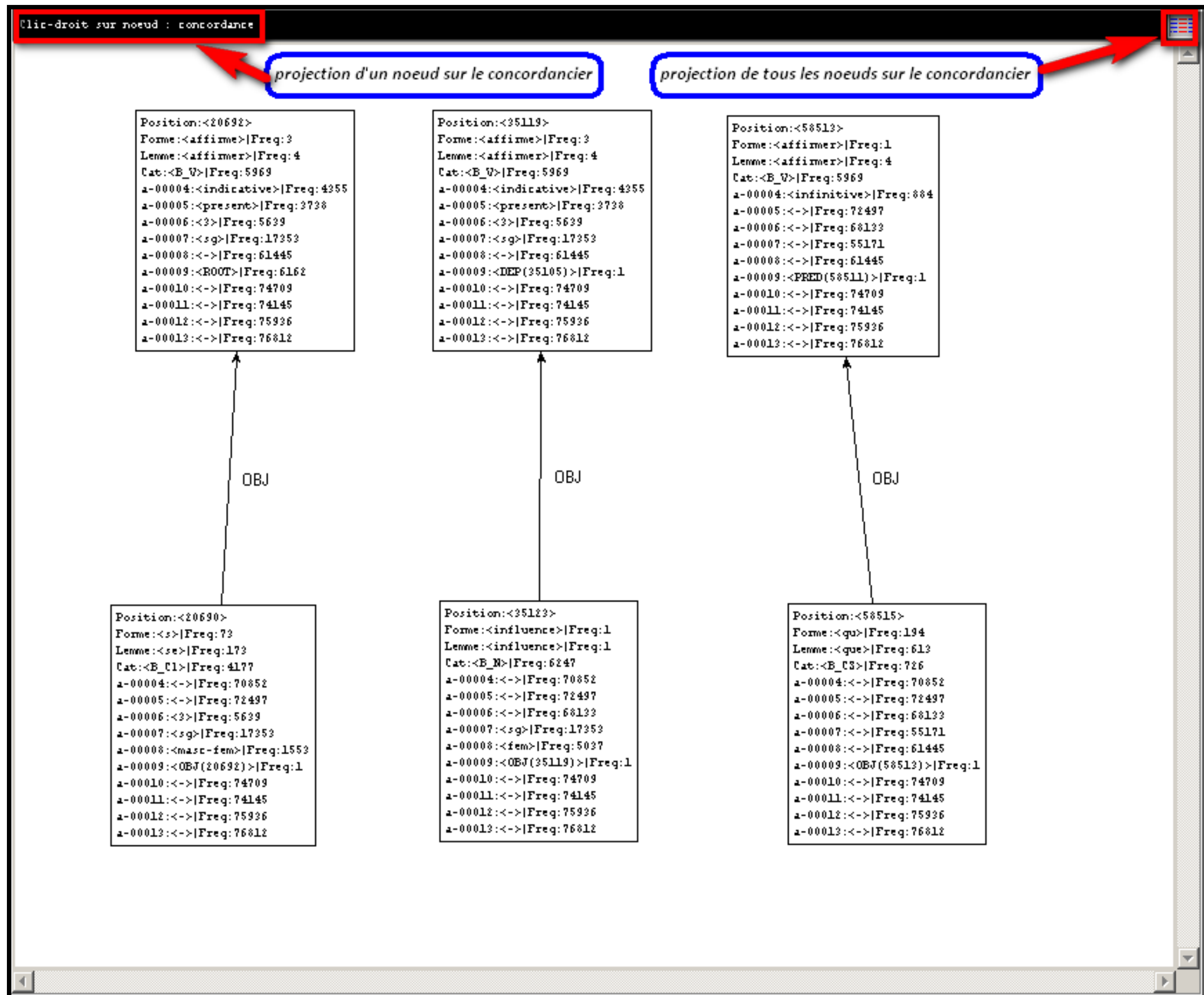


Dans les 2 graphes précédents, l'affichage des nœuds est celui produit par défaut : affichage de la position de l'item sur la *Trame* et de sa forme graphique. On peut visualiser l'ensemble des annotations d'un nœud donné via le mode Aperçu disponible dans tous les éditeurs du Trameur (en passant la souris au-dessus de l'objet visé) : ci-dessus, l'item en position 9899 (lemme : penser) est mis au jour.

Dans la figure qui suit, l'affichage d'un nœud est directement rendu « complet » via l'option du même nom disponible et activable avant de charger un graphe.

4.2 Retour en contexte

Un graphe de relations de dépendance étant produit, chaque nœud du graphe peut-être projeté sur le concordancier (`clic-droit` sur le nœud). L'ensemble des nœuds peut aussi être projeté globalement sur le concordancier :



La concordance produite dans ce dernier cas a l'allure suivante :

```
------------------------------PARTIE{PARTIE=M2004}------------------------------
                 § un nouvel ordre international s'affirme peu à peu
------------------------------PARTIE{PARTIE=M2001}------------------------------
               qu'une nation affirme son influence qu'elle pèse dans
------------------------------PARTIE{PARTIE=D2013}------------------------------
                  Rachida Dati pour oser affirmer qu'il ne s'agit
```

Les sources de la relation sont coloriées en rouge (pôle de la concordance), les cibles en vert (*i.e* ici le verbe est en vert, son objet en rouge).

On peut aussi varier l'affichage des annotations dans la concordance en matérialisant par exemple la catégorie de chaque item :

```
------------------------------PARTIE{PARTIE=M2004}------------------------------
 $ un-B_D  nouvel-B_Adj  ordre-B_N  international-B_Adj  s-B_Cl '-I_Cl affirme-B_V  peu-B_Adv  à-I_Adv  peu-I_Adv
------------------------------PARTIE{PARTIE=M2001}------------------------------
qu-B_Qu '-I_Qu une-B_D  nation-B_N  affirme-B_V  son-B_D  influence-B_N  qu-B_Qu '-I_Qu elle-B_Cl  pèse-B_V  dans-B_Pre
------------------------------PARTIE{PARTIE=D2013}------------------------------
 Rachida-B_N  Dati-B_N  pour-B_Pre  oser-B_V  affirmer-B_V  qu-B_CS '-I_CS il-B_Cl  ne-B_Cl  s-B_Cl '-I_Cl agit-B_V
```

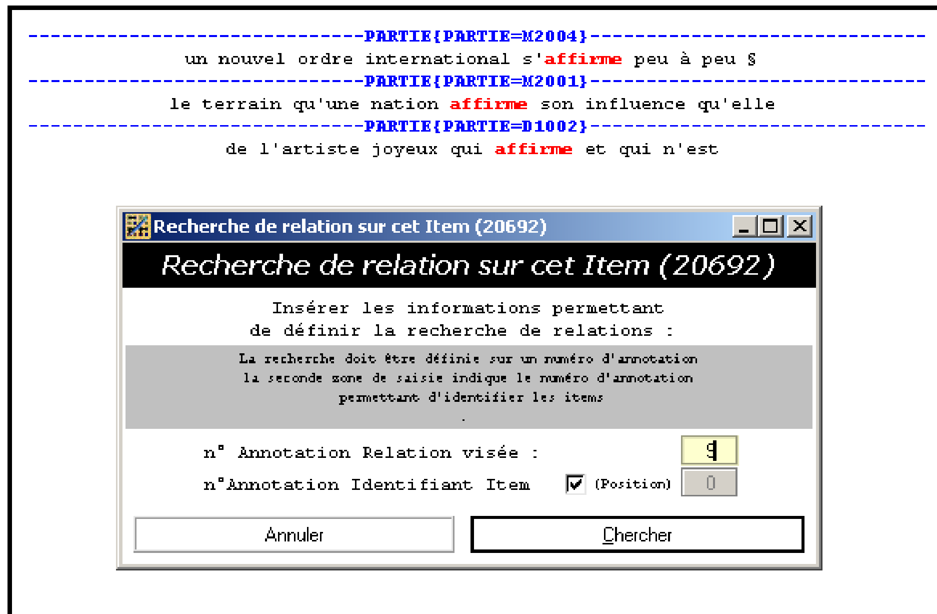
Toutes les zones d'édition du *Trameur* permettent de mettre au jour les annotations de la *Trame* (cf documentation en ligne, partie « Marquage des annotations de la *Trame* »). Si on considère la figure suivante, elle présente un extrait de la concordance construite à partir du graphe ayant permis d'extraire la relation OBJ (cf graphe infra) :

```
--------------------------------PARTIE{PARTIE=M2006 }--------------------------------
$ de la crise aux Antilles il ne devrait pas en être question au sommet social de l'Élysée cet
savamment minutée $ alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et
alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va
d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va se passer
passer avec vous Jean-François Achilli $ alors passons maintenant au détail des mesures discutées et aux attentes des syndicats
des mesures de justice je cite pour les salariés touchés par la crise économique $ mais Sara Ghibaudo il ne
vous d'aujourd'hui les Français se montrent très sévères à l'égard de la politique économique du gouvernement qu'
des voitures $ bonjour Sébastien Paour $ vous vous trouvez au Gosier $ où en est la situation $ merci
$ tout le monde est maintenant en tout cas suspendu à ce que Nicolas Sarkozy annoncera demain lors de sa
a promulgué son plan de relance adopté ce week-end par le congrès $ sept cent quatre-vingt-sept milliards
$ sept cent quatre-vingt-sept milliards de dollars destinés à sauver ou créer plus de trois millions et demi
à ralentir les saisies immobilières devrait être annoncé aujourd'hui par le président américain $ cela concerne plusieurs millions d'
Motors et Chrysler ont présenté hier leur plan de restructuration à l'administration Obama $ ils demandent vingt-deux milliards
toujours en garde à vue $ les enquêteurs cherchent maintenant à savoir si elle a été complice dans cette évasion
a été complice dans cette évasion $ elle se trouvait dans le parloir de la prison au moment où les
la foire d'empoigne $ la tension est encore montée d'un cran hier $ les avocats du berger corse
procès en attendant que la cour d'assise se prononce sur leur demande de supplément d'information $ lundi le
d~ ce sont des procédés terroristes $ il le dit à Laurent Doulsant $ direction maintenant Barcelone pour un congrès
Laurent Doulsant $ direction maintenant Barcelone pour un congrès consacré au téléphone mobile en crise lui aussi $ pour la
nul zéro partout pour Lille au Mans $ Valenciennes sort de la zone de relégation après sa victoire sur Caen
sa victoire sur Caen deux zéro $ Nice s'impose à Nancy deux à un $
--------------------------------PARTIE{PARTIE=M0023 }--------------------------------
voit euh qui voit ça passer $ et donc elle se dit eh bah elle peut récupérer elle pourra récupérer
rencontrent $ et et donc elle elle tombe $ ils se & $ ouais il y a un accident quoi
ouais il y a un accident quoi $ et ils se & ils tombent tous les deux $ et c'
boulanger voit que la baguette a disparu $ et il se dit que & $ ouais moi je pense qu'
va & $ alors le euh euh le boulanger dit au policier qu'en fait c'est la fille $
```

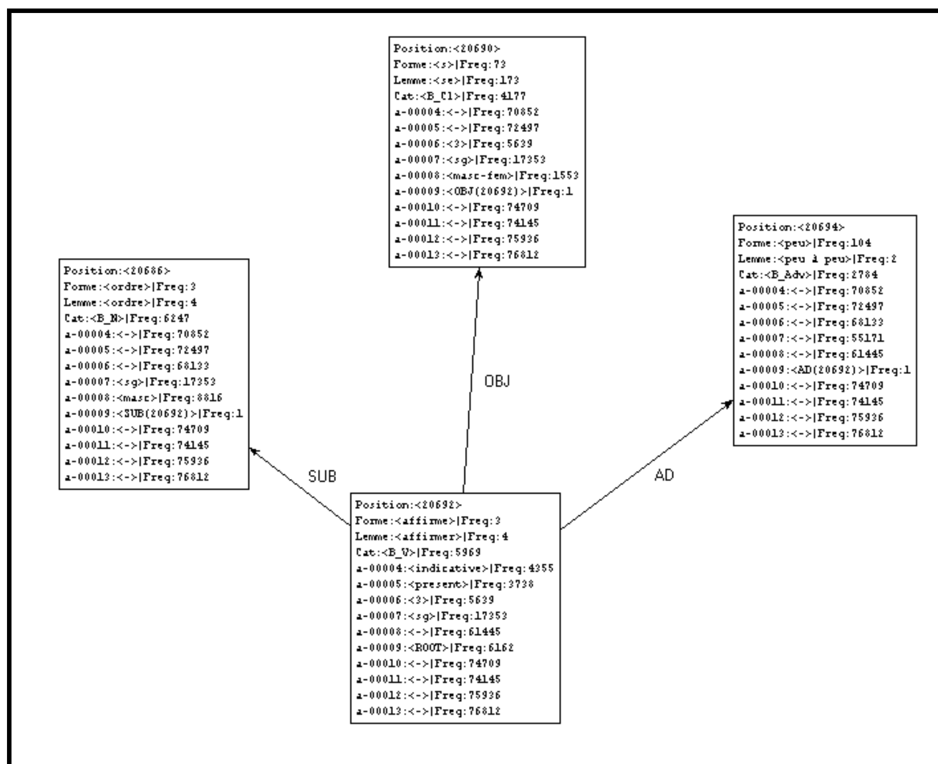
Le marquage de certains items est possible en sélectionnant les annotations à mettre au jour :

4.3 Recherche de dépendance en contexte

La recherche de dépendance peut aussi être réalisée en contexte, par exemple sur le concordancier. Si on considère la concordance suivante (en haut de la figure) :



Le raccourci clavier `Ctrl-Clic-droit` sur un item de la concordance (ici l'item visé est la première occurrence de la forme graphique « affirme ») permet de rechercher toutes les relations de dépendance pointant sur cet item. On commence par indiquer où chercher les relations de dépendance (numéro d'annotation portant ce type d'information : ici le n°9) et comment sont indexer les items sur la trame (ici par leur position). La recherche conduit à la production d'un graphe mettant au jour toutes les relations sur l'item visé :



Le graphe donne à voir les 3 relations pointant sur l'item.

4.4 Rechercher dans un graphe de dépendance

On présente tout d'abord le résultat produit par la requête suivante « recherche de la relation OBL » :

Sélection de relation à afficher

Sélection des relations à afficher

Insérer les informations (avec REGEXP)
permettant de définir la relation visée :

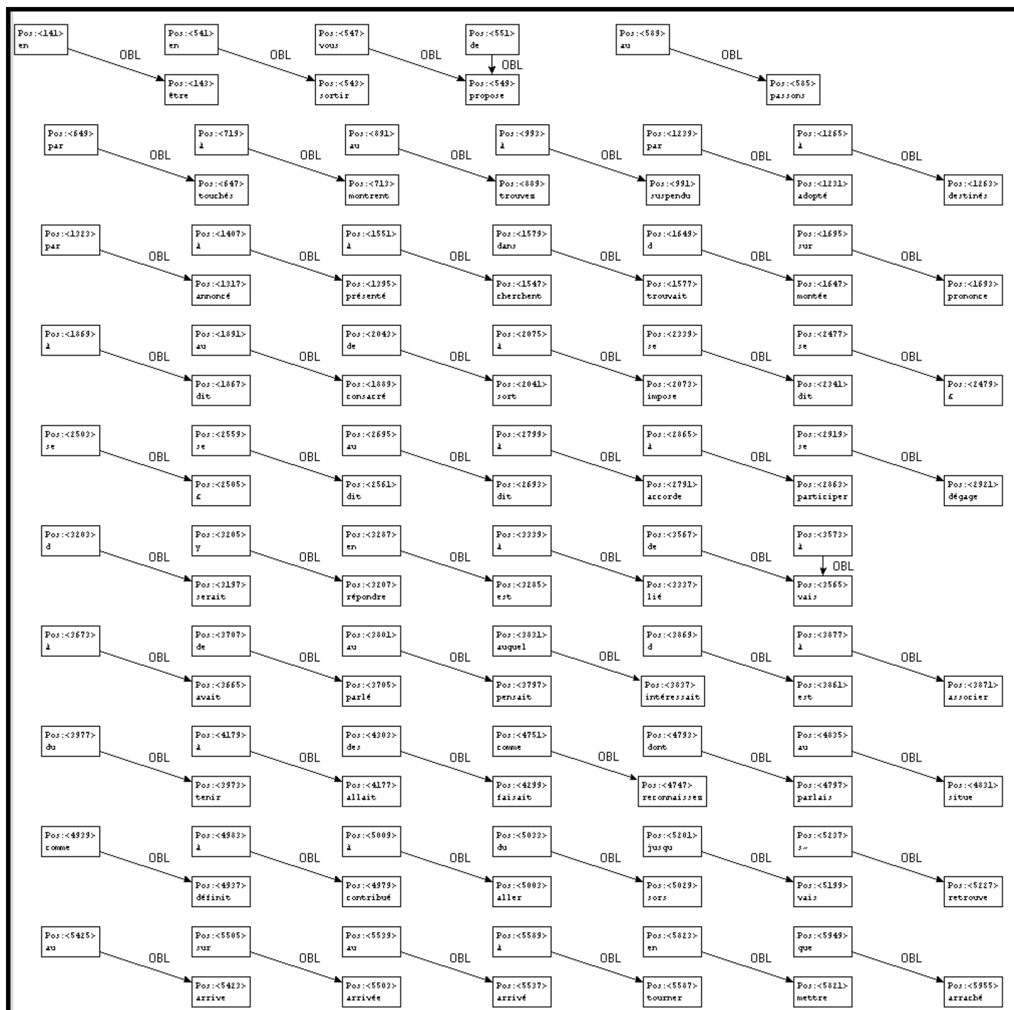
La relation visée doit être définie dans la première zone de saisie
par le nom de la relation (majuscule) et par son numéro d'annotation
correspondant, puis dans la seconde zone de saisie pour indiquer le
numéro d'annotation permettant d'identifier les items
Ex : Relation visée : SUJ / Annotation des relations : 7
n°Annotation des identifiants des items : 4
Le bouton "Options" permet d'afficher d'autres zones de saisie
permettant de filtrer les items associées aux relations.

Relation visée : OBL 9

n°Annotation Identifiant Item ☑ (Position) 0

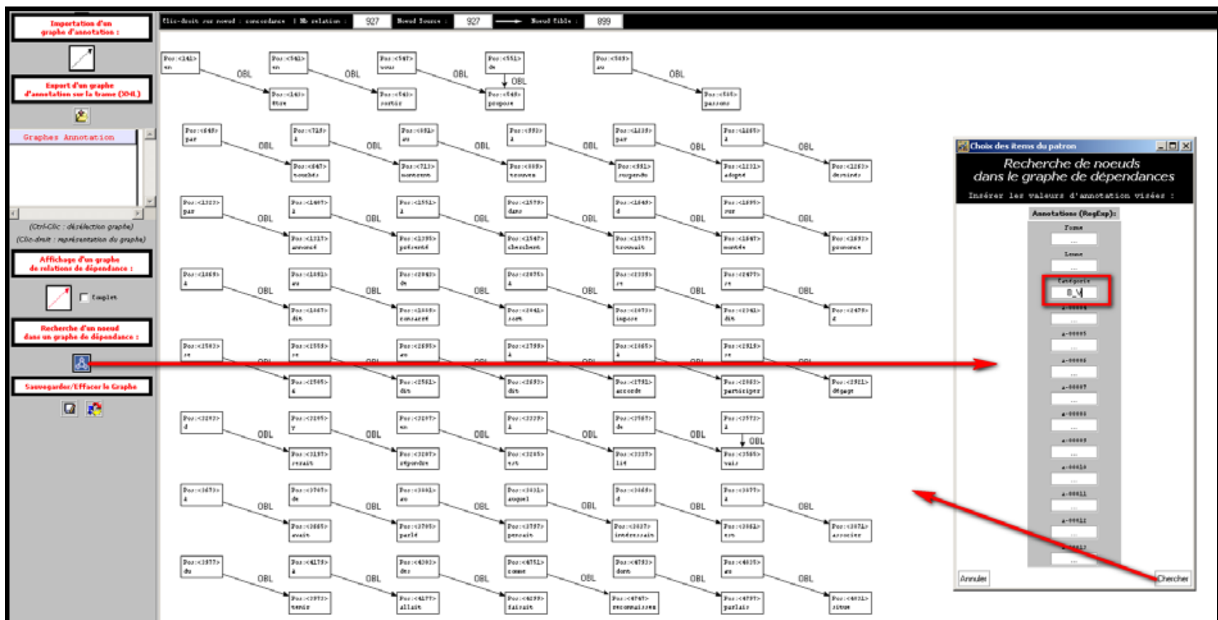
Options Annuler Chercher

Le résultat produit a l'allure suivante :

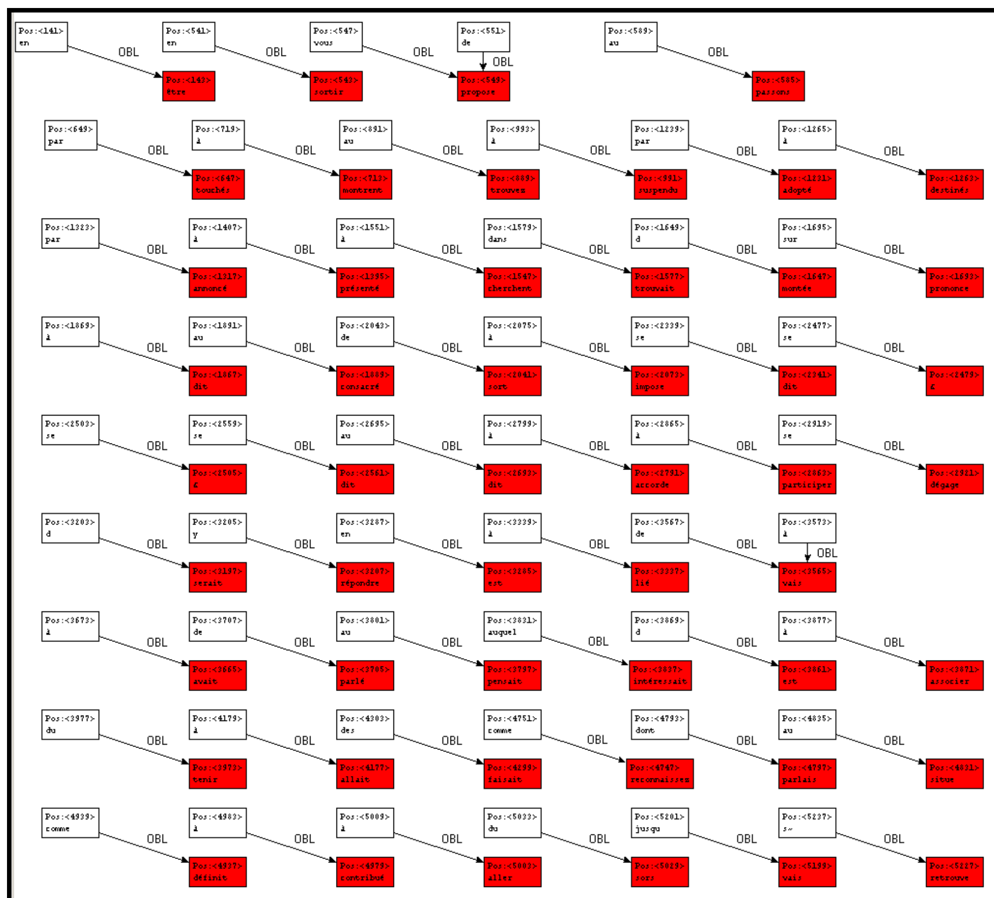


On peut mettre au jour dans ce graphe certains nœuds en filtrant leurs annotations.

Le processus de filtrage des nœuds du graphe permet de sélectionner certains nœuds sur la base des différentes annotations disponibles. Dans l'exemple suivant, on veut mettre au jour les nœuds de catégorie B_V :



Le résultat produit a l'allure suivante :



Dans la figure qui suit, 3 filtrages successifs sont réalisés pour marquer successivement les nœuds ayant pour catégorie B_V, puis B_N et enfin B_Pre ; chaque requête est précédée par la modification de la couleur à utiliser pour le marquage des nœuds visés (verbe en bleu, nom en vert et préposition en orange) :



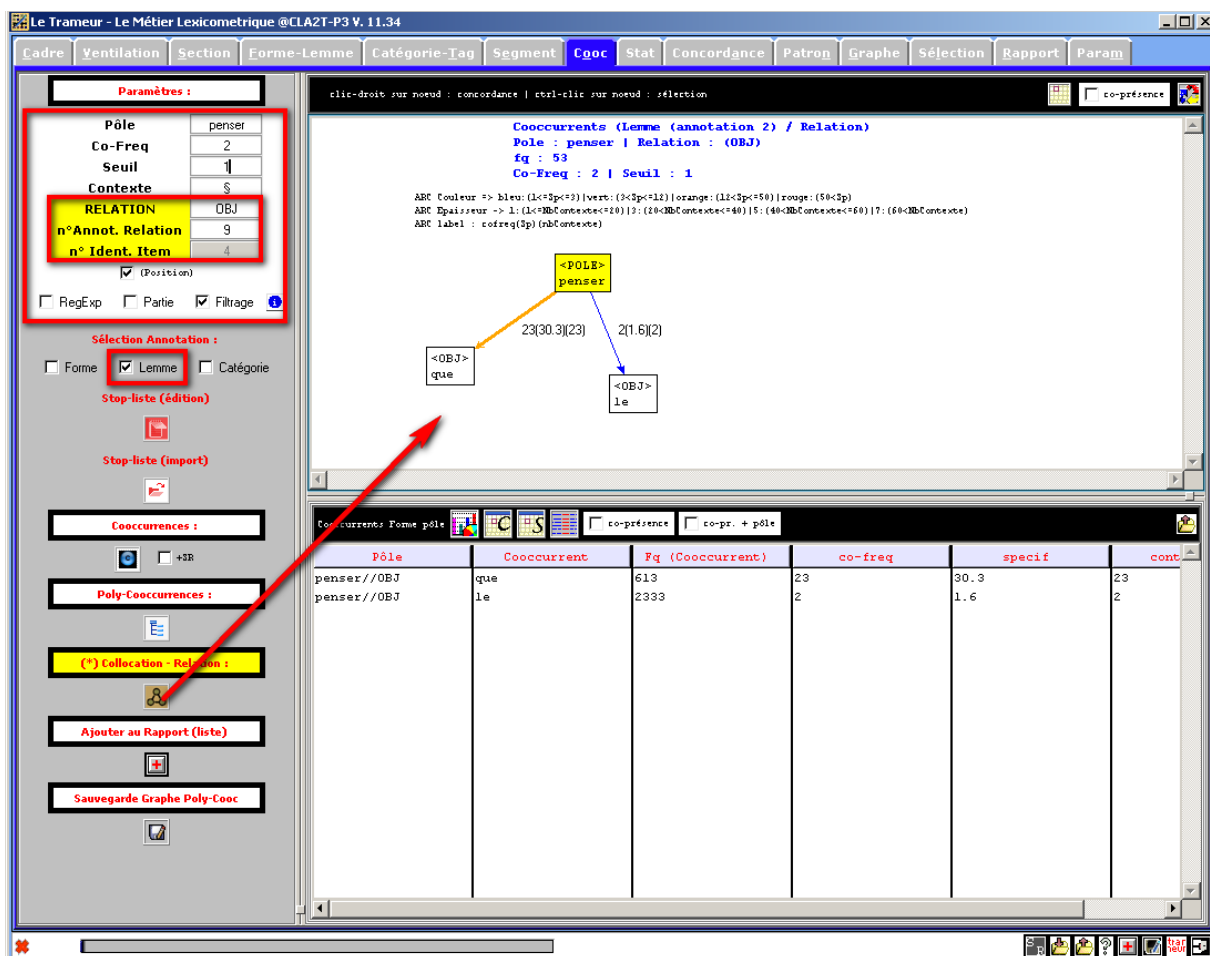
5. Recherche de collocation : spécificités sur relation

Une version particulière du module de calcul des cooccurrences permet de prendre en compte les relations entre les items de la Trame :

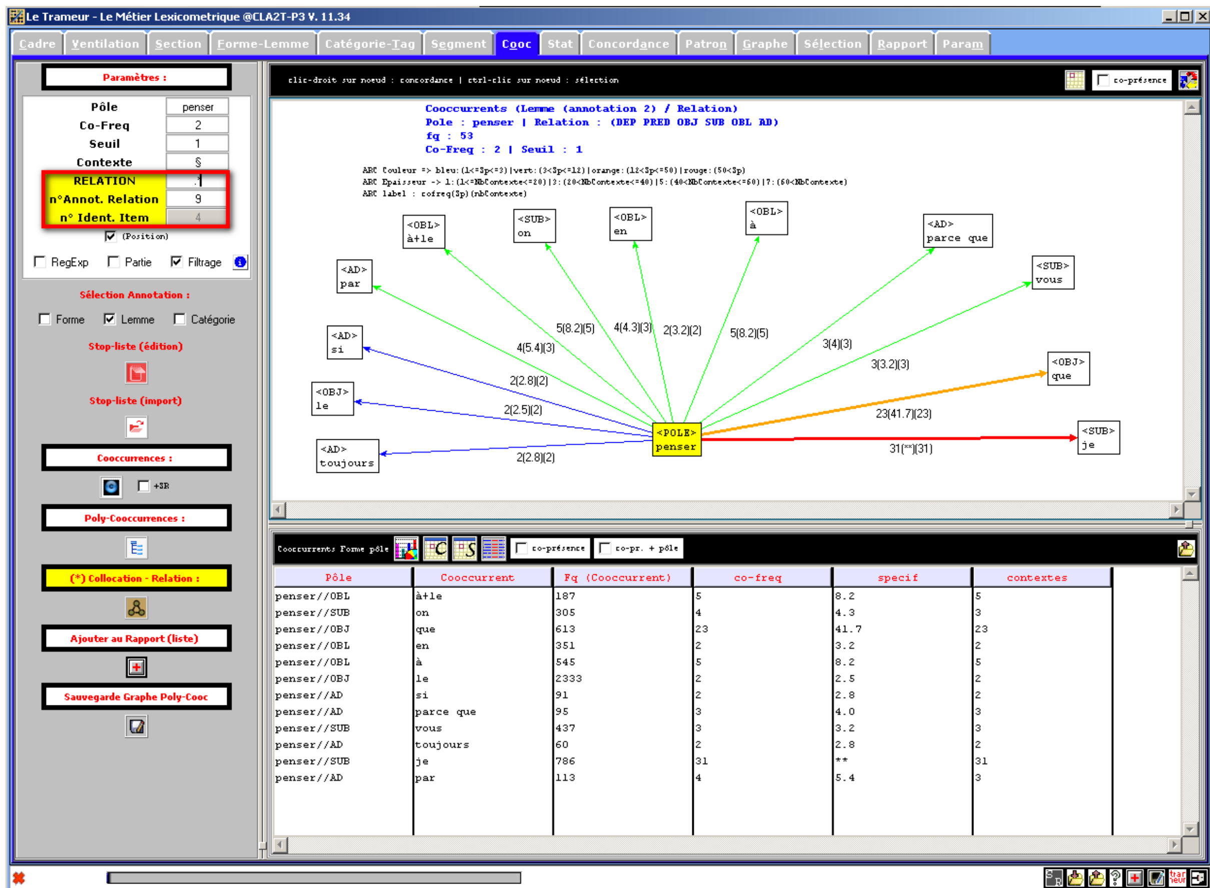
Etant donné une forme pôle, la recherche de ses cooccurrents peut être contrainte par la prise en compte d'une relation entre ce pôle est ses candidats cooccurrents. Pour un item X donné (le pôle), on s'intéresse aux items Y cooccurrents de X et en relation REL avec X (les collocatifs de X) :

X -> REL -> Y

La figure suivante présente les cooccurrents du lemme penser en contraignant ses cooccurrents à être en position d'objet :



La relation entre le pôle et ses cooccurrents peut être « non déterminée » par avance, dans la figure suivante, la relation est exprimée sous la forme : .* (regexp) ; elle vise donc toutes les relations du type : `penser-OBJ-y`, `penser-SUB-y` etc.



Une fois le graphe construit, un `clic-droit` sur un des noeuds montrent les contextes dans lesquels la relation se réalise (*i.e* les contextes utilisés par le calcul).

Dans cet exemple, 3 contextes dans lesquels *vous* est sujet de *penser* (*cf* arc `penser -> vous`) ont été mis au jour, un `clic-droit` sur le nœud *vous* les montrent dans le concordancier (les 2 items de la relation y sont colorés automatiquement : *vous* en rouge (pôle de la concordance) et *penser* en vert).

```
-----------------------------PARTIE{PARTIE=D2009 }-----------------------------
un mot DELIM ou bien vous le penser vraiment DELIM
-----------------------------PARTIE{PARTIE=D0006 }-----------------------------
dans le quartier de Paris vous en penser quoi ici
-----------------------------PARTIE{PARTIE=D2011 }-----------------------------
DELIM que^êtreUNKNOWNce que vous en penser de le
Position:<64987>
Forme:<vous>|Freq:441
Lemme:<vous>|Freq:437
Cat:<B_Cl>|Freq:4177
a-00004:<->|Freq:71023
a-00005:<->|Freq:72668
a-00006:<2>|Freq:988
a-00007:<pl>|Freq:3992
a-00008:<->|Freq:61616
a-00009:<SUB(64991)>|Freq:1
a-00010:<->|Freq:74880
a-00011:<->|Freq:74316
a-00012:<->|Freq:76107
a-00013:<->|Freq:76983
```